

[Home Page](#) • [Local Home Page](#) • [Text](#) • [Assignments](#)   [Schedule](#)   [Links](#)

# Bioinformatics for protein sequence, structure and function



## Homology modeling of protein structures

Homology modeling involves taking a known sequence with an unknown structure and mapping it against a known structure of one or several similar (homologous) proteins. It would be expected that two proteins of similar origin and function would have reasonable sequence similarity. One can then attempt to say that a conserved region of sequence that is, say for example, an alpha-helix in the known protein, is probably the same in the unknown's sequence.

### Basic procedures utilized in Most homology modeling:

1. Obtaining the sequence(s) and checking with an alignment program, such as BLAST or FASTA to ensure relatedness and homology
2. Finding PDB data for known structures (NMR data can also help, but a 3-D coordinate map is necessary)
3. Multiple sequence alignments. (if more than one known is involved, sometimes the knowns are aligned together, then the unknown sequence aligned with the group; this helps ensure better domain conservation) Analysis of alignments; gap deletions and additions; secondary structure weighting
4. Structure calculation
5. Model refinement

The main difference between the different comparative modeling methods is in how the 3D model is calculated from a given alignment. The original and still the most widely used method is modeling by rigid body assembly. The method constructs the model from a few rigid bodies that include core regions, loops and side-chains, all of which are obtained from dissecting related structures. The assembly of the model involves calculating a framework, which is defined as the average of the template atoms in the conserved regions of the fold, and then fitting the rigid bodies on the framework. Another family of methods, modeling by segment matching, relies on approximate positions of conserved atoms from the templates to calculate the coordinates of other atoms. This is achieved by the use of a database of short segments of protein structure, energy or geometry rules, or some combination of these criteria. The third group of methods, modeling by satisfaction of spatial restraints, uses either distance geometry or optimization techniques to satisfy spatial restraints obtained from the alignment of the target sequence with similar templates of known structure. Some available software packages for comparative modeling are listed in Table 1. In addition to the methods for modeling the whole fold, numerous other techniques for predicting loops and side-chains on a given backbone have also been described. These methods can often be used in combination with each other and with comparative modeling techniques.

### The modeller approach

Please read [this](#) paper

## The ICM approach

ICM is more of a classical molecular modeling program than modeller. What separates it from other modeling approaches are (1) the energy function and (2) the representation of the proteins. The energy function contains terms to rapidly calculate a (good) approximation of entropy and hydrophobic effects. Below is the methods for ICM homology modeling outlined.

1. check and adjust you sequence-template alignment with the molecular graphics
2. build initial model by threading your sequence onto the template and patching insertions and deletions with results of the built-in database fragment search
3. find the lowest energy loop conformations by the global PBMC-loop optimization in all-atom representation, in the soft environment and with inclusion of solvation energy and entropy.
4. predict conformations of chain ends missing in the template.
5. predict side-chain conformations by fast continuous PBMC free energy optimization.

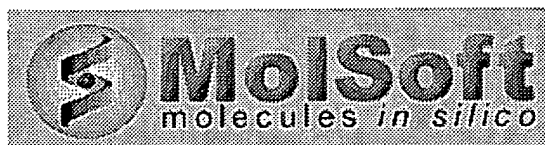
---

Arne Elofsson

Last modified: Wed Oct 27 15:47:19 CEST 1999

---

<b>Arne Elofsson</b>	<b>Tel: +46-(0)8/161553</b>
<b>Stockholm Bioinformatics Center,</b>	<b>Fax: +46-(0)8/158057</b>
	<b>Hem: +46-(0)8/6413158</b>
<b>Department of Biochemistry,</b>	
<b>Arrheniuslaboratoriet</b>	<b>Email: <a href="mailto:arne@sbc.su.se">arne@sbc.su.se</a></b>
<b>Stockholms Universitet</b>	<b>WWW: <a href="http://www.sbc.su.se/~arne/">/~arne/</a></b>
<b>10691 Stockholm, Sweden</b>	



Search

GO

**Info**

[Home](#)  
[About us](#)  
[Technology](#)  
[Contact us](#)

**News**

[Business](#)  
[ICM](#)

**Products**

[ICM-Main](#)  
[ICM-Pro](#)  
[ICM-Lite](#)  
[XPDB](#)

**Services**

[Overview](#)  
[ICM VLS](#)

**Online Tools**

[PDB-Viewer](#)  
[2D to 3D Converter](#)

**Downloads**

[ICM full version](#)  
[ICM lite version](#)

**Support**

[ICM Online Manual](#)  
[ICM Support E-mail](#)

**References**

- Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2002). Screened charge electrostatic model in protein-protein docking simulations. *Pac Symp Biocomput.* 2002;:552-63
- Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2002). Soft Protein-Protein Docking in Internal Coordinates. *Protein Science* 11:280-91
- Totrov, M., and Abagyan, R. (2001). Protein-ligand docking as an energy optimization problem. *Drug-Receptor Thermodynamics: Introduction and Applications*. Editor: R.B. Raffa, John Wiley & Sons, RV, 603-624.
- Totrov, M., and Abagyan, R. (2001). Rapid boundary element solvation electrostatics calculations in folding simulations: Successful folding of a 23-residue peptide. *Biopolymers* 2001;60(2):124-33.
- Schapira, M., Raaka, B. M., Samuels, H. H., and Abagyan, R. (2001). In Silico Discovery of novel Retinoic Acid Receptor Agonist Structures. *BMC Structural Biology Journal* 2001;1 (1):1
- Norledge, B.V., Lambeir, A.M., Abagyan, R.A., Rottmann, A., Fernandez, A.M., Filimonov, V., Peter, M.G., and Wierenga, R.K. (2001). Modeling, mutagenesis, and structural studies on the fully conserved phosphateloop (loop 8) of triosephosphate isomerase: toward a new substrate specificity. *Proteins* Feb 15;42(3), 383-9
- Abagyan, R. and Totrov, M. (2001). High-throughput Docking for Lead Generation. *Current Opinion in Chemical Biology*. Aug;5(4):375-82
- Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, James TL.(2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J Comput Aided Mol Des.* 14(6):593-610.
- Tomko, R.P., Totrov, M., Abagyan, R. and Philipson, L. (2000). Expression of the Adenovirus Receptor and Its Interaction with the Fiber Knob. *Experimental Cell Research*, Feb 25;255(1), 47-55.
- Schapira, M., Raaka, B.M., Samuels, H. H. and Abagyan, R. (2000). Rational design of novel nuclear hormone receptor antagonists. *PNAS* 97 (3), 1008-1013.
- Jin, E., Katritch, V., Olson, W.K., Kharatisvili, M., Abagyan, R., Pilch, D.S. (2000) Aminoglycoside binding in the major groove of duplex RNA: the thermodynamic and electrostatic forces that govern recognition. *J. Mol. Biol.* Apr 21 298 (1), 95-110.
- Kelly, P.D., Chu, F., Wood, I.G., Ngo-Hazelett, P., Cardozo, T., Huang, H., Kimm, F., Liao, L., Yan, Y.L., Zhou, Y., Johnson, S.L., Abagyan, R., Schier, A.F., Postlethwait, J.H.,

- Talbot, W.S. (2000) Genetic linkage mapping of zebrafish genes and ESTs. *Genome Res. Apr. 10 (4)*, 558-567.
- Gantt, S., Persson, C., Rose, K., Birkett, A.J., Abagyan, A. and Nussenzweig, V. (2000) Antibodies against TRAP do not inhibit Plasmodium sporozoite infectivity in vivo. *Infection and Immunity Jun; 68(6)*, 3667-3673.
  - Stigler, R.-D., Hoffmann, B, Abagyan, R. and Schneider-Mergener, J. (1999) Soft docking an L and a D peptide to an anticholera toxin antibody using internal coordinate mechanics. *Structure, 7*, 663-670.
  - Zhang HF, Yu J, Chen S, Morgan BP, Abagyan R, Tomlinson S. (1999) Identification of the individual residues that determine human CD59 species selective activity. *J Biol Chem 1999 Apr 16;274(16):10969-74*
  - Schapira, M., Totrov, M. and Abagyan, R. Prediction of the binding energy for small molecules, peptides and proteins. *J Mol Recognit 12*, 177-190 (1999).
  - Li, D., Desai-Yajnik, V., Lo, E., Schapira, M., Abagyan, R. and Samuels, H.H. NRIF3 Is a Novel Coactivator Mediating Functional Specificity of Nuclear Hormone Receptors. *Mol Cell Biol 19*, 7191-7202 (1999) .
  - Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, Talbot WSA. (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res 1999 Apr;9(4):334-47*
  - Zhou Y, Abagyan R. (1999) Efficient Stochastic Global Optimization for Protein Structure Prediction. *Rigidity Theory and Applications*, Plenum Publishing, New York, 513-522.
  - Zhou Y, Abagyan R. (1998) How and why phosphotyrosine-containing peptides bind to the SH2 and PTB domains. *Folding & Design.*, 3, 513-522.
  - Cardozo, T.J., Abagyan, R. (1998) Molecular Modeling of the Domain Shared Between CED-4 and its Mammalian Homologue Apaf-1: A Structural Relationship to the G-proteins. *J. Mol. Model.*, 4, 83-93.
  - Maiorov, V. & Abagyan, R. (1998) Energy strain in three-dimensional protein structures *Folding & Design.*, 3, 259-269.
  - Patel, I.R., Attur M.G., Patel R.N., Stuchin S.A., Abagyan R.A., Abramson S.B., Amin A.R. (1998) TNF-alpha convertase enzyme from human arthritis-affected cartilage: isolation of cDNA by differential display, expression of the active enzyme, and regulation of TNF-alpha. *J. Immunol.* 160, 4570-4579.
  - Srivastava S, Osten P, Vilim FS, Khatri L, Inman G, States B, Daly C, DeSouza S, Abagyan R, Valtschanoff JG, Weinberg RJ, Ziff EB., (1998) Novel anchorage of GluR2/3 to the postsynaptic density by the AMPA receptor-binding protein ABP. *Neuron 21(3):581-91*.
  - Totrov, M. & Abagyan, R. (1997) Flexible protein-ligand docking by global energy optimization in internal coordinates

- Proteins*, Suppl. 1, 215-220 .
- Abagyan, R., Batalov, S., Cardozo, T., Totrov, M., Webber, J. and Zhou, Y. (1997) Homology Modeling with Internal Coordinate Mechanics: Deformation Zone Mapping and Improvements of Models via Conformational Search. *Proteins: Struct. Funct. & Gen.*, Suppl. 1, 29-37.
  - Abagyan, R.A. (1997) Protein structure prediction by global energy optimization. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Volume 3, Eds. van Gunsteren, W.F. et al., ESCOM Science Publishers BV, Leiden.
  - Mathieu, M., Modis, Y., Zeelen, J.P., Engel, C.K., Abagyan, R.A., Ahlberg, A., Rasmussen, B., Lamzin, V.S., Kunau, W.H., Wierenga, R.K. (1997) The 1.8Å crystall structure of the dimeric peroxisomal 3-ketoacyl-coA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *J. Mol. Biol.*, 273, 714-728.
  - Rashin, A.A., Rashin, B.H., Rashin A., Abagyan, R. (1997) Evaluating the energetics of empty cavities and internal mutations in proteins. *Protein Science*, 6, 2143-2158.
  - Koonin, E.V. & Abagyan, R.A. (1997) TSG101, a gene mutated in human breast cancer, encodes a ubiquitin-conjugating enzyme homolog with replaced active cysteine. *Nature Genetics*, 16, 330-331.
  - Yu J, Dong S, Rushmere NK, Morgan BP, Abagyan R, Tomlinson S., (1997) Mapping the regions of the complement inhibitor CD59 responsible for its species selective activity. *Biochemistry* 36, 9423-9428.
  - Yu, J., Abagyan, R.A., Dong, S., Gilbert, A., Nussenzweig, V., Tomlinson, S. (1997) Mapping the active site of CD59 *J. Exp. Med.* 185, 745-753.
  - Abagyan R.A. & Batalov, S. (1997) Do aligned sequences share the same fold? *J. Mol. Biol.* 273, 355-368.
  - Abagyan, R.A. & Totrov, M.M. (1997) Contact Area Difference (CAD): A robust measure to evaluate accuracy of protein models. *J. Mol. Biol.* 286, 678-685.
  - N.Thanki, J.Ph.Zeelen, M. Mathieu, R.Jaenicke, R.A.Abagyan, R.K.Wierenga and W.Schliebs (1997) Protein engineering with monomeric triosephosphate isomerase (monoTIM): the modelling and structure verification of a seven residue loop. *Protein Eng.*, 10, 159-167.
  - Maiorov, V.N., Abagyan, R.A. (1997) A new method for modeling large-scale rearrangements of protein domains. *Proteins* 27, 410-424.
  - Strynadka, N.C., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B.K., Kuntz, I.D., Abagyan, R.A., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson A., Duncan, B., Rao, M., Jackson, R. Sternberg, M., James, M.N. (1996) Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat.*

- Struct. Biol.*, 3, 233-239.
- Chalikian, T.V., Totrov, M.M., Abagyan, R.A., Breslauer, K.J. (1996) The hydration of globular proteins as derived from volume and compressibility measurements: cross correlating thermodynamic and structural data. *J. Mol. Biol.*, 260, 588-603.
  - Totrov, M.M., and Abagyan, R.A. (1996) The contour-buildup algorithm to calculate the analytical molecular surface. *J. Struct. Biol.*, 115, 1-6.
  - Borchert, T.V., Kishan, K.V., Zeelen, J.P., Schliebs, W., Thanki, N., Abagyan, R., Jaenicke, R. & Wierenga, R.K. (1995) Three new crystal structures of point mutation variants of monoTIM: conformational flexibility of loop-1, loop-4 and loop-8. *Structure*, 3, 669-679.
  - Cardozo, T., Totrov, M.M. & Abagyan, R.A (1995) Homology modeling by the ICM method. *Proteins*, 23, 403-414.
  - Abagyan, R.A., Frishman, D., & Argos, P. (1994) Recognition of distantly related proteins through energy calculations. *Proteins* 19, 132-140.
  - Abagyan, R.A. & Totrov, M.M. (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235, 983-1002.
  - Abagyan, R.A., Totrov, M.M. & Kuznetsov, D.N. (1994) ICM - a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. *J.Comp.Chem.* 15, 488-506.
  - Borchert, T.V., Abagyan, R.A., Jaenicke, R., & Wierenga, R.K. (1994) Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci USA* 91, 1515-1518.
  - Totrov, M.M. & Abagyan, R.A. (1994) Efficient parallelization of the energy, surface and derivative calculations for internal coordinate mechanics. *J. Comp. Chem* 15, 1105-1112.
  - Totrov, M.M. & Abagyan, R.A. (1994) Detailed *ab initio* prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nat. Struct. Biol.* 1, 259-263.
  - Abagyan, R.A. (1993) Towards protein folding by global energy optimization. *FEBS Letters* 325, 17-22.
  - Borchert, T.V., Abagyan, R.A., Kishan, K.V.R., Zeelen, J.Ph., & Wierenga, R.K. (1994) The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: the correct modelling of an eight-residue loop. *Structure* 1, 205-213.
  - Eisenmenger, F., Argos, P., & Abagyan, R.A. (1993) A method to configure protein side-chains from the main-chain trace in homology modelling. *J. Mol. Biol.* 231, 849-860.
  - Abagyan, R.A. & Argos, P. (1992) Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* 225, 519-532.

---

### The main description of the ICM method

- Abagyan, R.A. and Totrov, M.M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235, 983-1002.
- Abagyan, R.A., Totrov, M.M., and Kuznetsov, D.N. (1994). ICM - a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J.Comp.Chem.* 15, 488-506.

---

### Other related publications

- Abagyan, R.A. and Maiorov, V.N. (1988). A simple quantitative representation of polypeptide chain folds: comparison of protein tertiary structures. *J. Biomol. Struct. Dyn.* 5, 1267-1279.
- Abagyan, R.A. and Mazur, A.K. (1989). New methodology for computer-aided modeling of biomolecular structure and dynamics. 2. Local deformations and cycles. *J. Biomol. Struct. Dyn.* 6, 833-845.
- Abagyan, R.A. and Argos, P. (1992). Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* 225, 519-532.
- Borchert, T.V., Abagyan, R.A., Kishan, K.V.R., Zeelen, J.Ph., & Wierenga, R.K. (1994). The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: the correct modeling of an eight-residue loop. *Structure* 1, 205-213.
- Borchert, T.V., Abagyan, R.A., Jaenicke, R., and Wierenga, R.K. (1994). Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci USA* 91, 1515-1518.
- Eisenmenger, F., Argos, P., and Abagyan, R.A. (1993). A method to configure protein side-chains from the main-chain trace in homology modeling. *J. Mol. Biol.* 231, 849-860.
- Mazur, A.K. and Abagyan, R.A. (1989). New methodology for computer-aided modeling of biomolecular structure and dynamics. 1. Non-cyclic structures. *J. Biomol. Struct. Dyn.* 6, 815-832.
- Totrov, M.M. and Abagyan, R.A. (1994). Efficient parallelization of the energy, surface and derivative calculations for internal coordinate mechanics. *J. Comp. Chem* 15, 1105-1112.
- Totrov, M.M. and Abagyan, R.A. (1994). Detailed *ab initio* prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nature Struct. Biol.* 1, 259-263.
- Totrov, M.M., and Abagyan, R.A. (1996). The contour-buildup

algorithm to calculate the analytical molecular surface. *J. Struct. Biol.* 115, 1-6.

- Abagyan, R.A., and Batalov, S. (1997) Do aligned sequences share the same fold? *J. Mol. Biol.*, 273, 1, 355-368.
- Abagyan, R. and Totrov, M. (1997) Contact Area Difference (CAD): A robust measure to evaluate accuracy of protein models. *J. Mol. Biol.* 268, 678-285.

---

### Applications of the ICM program

- Cardozo, T., Totrov, M., and Abagyan, R. (1995) Homology modeling by the ICM method. *Proteins* 23, 403-414 .
- Borchert, T.V., Kishan, K.V.R., Zeelen, J.Ph., Schliebs, W., Thanki, N., Abagyan, R.A., Jaenicke, R., and Wierenga, R.K. (1995 ) Three new crystal structures of point mutation variants of monoTIM: Conformational flexibility of loop-1, loop4 and loop8. *Structure* 3, 669-679.
- Strynadka, N.C.J., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B.K., Kuntz, I.D., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M., and James, M.N.G. (1996) Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nature Struct. Biol.* 3, 233-239.
- Chalikian, T.V., Totrov, M.M., Abagyan, R.A., Breslauer, K.J. (1996) The hydration of globular proteins as derived from volume and compressibility measurements: cross correlating thermodynamic and structural data. *J. Mol. Biol.* 260, 588-603.
- Thanki, N., Zeelen, J.Ph., Mathieu, M., Jaenicke, R., Abagyan, R.A., Wierenga, R.K., and W.Schliebs (1997) Protein engineering with monomeric triosephosphate isomerase (monoTIM): The modeling and structure verification of a seven residue loop. *Protein Eng.* 10, 159-167.
- Goodman, A.R., Cardozo, T., Abagyan, R.A., Altmeyer, A., Wisniewski, H.G., and Vilcek, J. (1996) Long pentraxins: an emerging group of proteins with diverse functions. *Cytokine Growth Factor Rev.* 7, 191-202.
- Maiorov, V.N. and Abagyan, R.A. (1997) A new method for modeling large-scale rearrangements of protein domains. *Proteins* 27, 410-424.
- Yu, J., Abagyan, R., Dong, S., Gilbert, A., Nusenzweig, V., & Tomlinson, S. (1997) Mapping of the Active Site of CD59. *J. Expt. Medicine* 185, 745-754.

[back to top.](#)